

# Cerca de relacions gèniques per un tòpic d'interès a bases de dades biomèdiques remotes

Sergi Farré i Casals

**Resum**— Els gens al expressar-se duen a terme les funcions de la cèl·lula. Així doncs modificant la seva expressió provoquen un canvi cel·lular que ens pot portar de l'estat sa o la patologia. Actualment hi ha bases de dades biomèdiques internacionals en les quals s'emmagatzema informació sobre si dos gens tenen cap tipus de relació. L'IBB, Centre d'Investigació de la UAB, té una línia de recerca per a l'anàlisi de matrius d'expressió gènica per Data Mining via web. Una d'aquestes eines web ens mostra les relacions d'expressió entre tots els gens d'una matriu d'expressió i també les relacions a partir de la informació de les bases de dades biomèdiques citades. No obstant, la cerca d'aquestes relacions té fortes limitacions, en part per la gran quantitat de dades que s'han de crear i en part perquè no permet dirigir la cerca d'aquestes relacions als tòpics d'interès per l'investigador/a. Combinant l'ús de bases de dades locals amb part de la informació remota, crides selectives a les bases de dades biomèdiques remotes per obtenir informació respecte els tòpics d'interès de l'usuari, i taules auxiliars i fitxers temporals, aconseguim: Proporcionar unes relacions gèniques centrades amb els tòpics d'interès dels investigadors/res i poder treballar amb la gran quantitat de dades contingudes a les bases de dades biomèdiques que emmagatzemen tot el coneixement biomèdic de la humanitat. Com a resultat disposem d'una eina molt potent, amb una interfície millorada i adaptada a les noves prestacions, i molta més capacitat d'interacció de l'usuari, tant a l'hora d'escollir el tipus de relació gènica que vol cercar, el tòpic d'interès, i el domini de gens a relacionar, com a l'hora d'estudiar les relacions gèniques proporcionades per l'aplicació.

**Paraules clau**—Gens, Microarray, Big Data, Web databases, NCBI, Pubmed, Data Mining, Relacions gèniques...

**Abstract**— Genes to express themselves carry out the functions of the cell. So changing its expression causes cell changes that can lead to disease or healthy state. There are currently international biomedical databases where information is stored if two genes have any relationship. IBB, the UAB Research Centre, has a line of research for the analysis of gene expression arrays for Data Mining web. One of these web tools shows the expression relationship between a gene array expression and also relationships based on the information of biomedical databases previously cited. However, finding these relationships has strong limitations in part by the large amount of data to be crossed, and partly because it allows direct search of these relationships to topics of interest to the researcher / a. Combining the use of local databases with information of remote calls selective biomedical databases to obtain information about the remote topics of interest to the user and tables and temporary files, manage: Provide gene relationships focused on topics of interest to researchers and to work with the large amount of data contained in the databases that store all biomedical knowledge of humanity. As a result we have a very powerful tool with an improved interface and adapted to the new features, and a lot more interaction from the user, both in choosing the type of relationship you want to search gene, the topic of interest and the domain of related genes, such as when to study genetic relationships provided by the application.

**Index Terms**— Genes, Microarray, Big Data, Web databases, NCBI, Pubmed, Data Mining, Genetic relations.



## 1 Introducció

La tecnologia de microarrays obté el nivell d'expressió d'un nombre gran de gens (1500-10000) sota un gran nombre de condicions mostrals diferents (100-1000). Aquestes matrius d'expressió ens permeten relacionar els gens que es coexpressen, és a dir que s'expressen alhora d'un determinat procés cel·lular. També ens permeten estudiar altres tipus de relacions d'expressió més complexes, com les relacions que ens mostren una alternança en l'activació i la desactivació dels gens. No obstant, les dades contingudes a la microarray no ens permeten relacionar els gens més enllà de la relació de les seves expressions.

El problema de determinar la funció d'una seqüència DNA pot ser afrontat des de diferents vessants, anàlisi de homologia, cerca de motius seqüencials, tècniques experimentals, etc.

Una altre aproximació al problema pot ser la extracció de coneixement a partir de l'anàlisi de microarrays [1-9]. La tecnologia de microarrays proporciona els nivells de l'expressió dels gens per diferents condicions experimentals. Al construir una microarray se'ns proporcionarà una matriu on les files representen els gens, les columnes els experiments i les cel·les, la expressió del determinat gen per a aquest experiment.

L'IBB [10], centre d'investigació de la Universitat Autònoma de Barcelona (UAB), té una línia d'investigació per l'anàlisi de microarrays [1-9]. En relació a aquesta línia d'investigació disposem d'un servidor a l'IBB (<http://revolutionresearch.uab.es>) [11] on podem trobar l'aplicació web PCOPGene [3] que permet estudiar i analitzar microarrays amb diverses eines desenvolupades al IBB. Aquestes eines desenvolupades a l'IBB per l'anàlisi

- E-mail de contacte: [sergiengineer@gmail.com](mailto:sergiengineer@gmail.com)
- Menció realitzada: Enginyeria del Software.
- Treball tutoritzat per: Mario Huerta (IBB) i Jordi González (CVC).
- Curs 2014/15

### 1.1 Estat de l'art

de microarrays han permès, per exemple, resoldre la paradoxa sobre els efectes contradictoris dels glucocorticoides (com la cortisona) en el desenvolupament dels tumors [5].

Aquesta aplicació web, PCOPGene [3], també permet el creuament dels resultats d'analitzar una microarray amb bases de dades biomèdiques remotes [12-16]. Una funcionalitat que resulta molt interessant pels investigadors i investigadores, ja que permet trobar relacions entre gens de la microarray més enllà de les relacions a nivell d'expressió, que són les que es poden extreure a partir de les dades de microarray.

Aquestes relacions gèniques poden ser obtingudes de diferents bases de dades d'àmbit internacional [12-14]:

- Bases de dades que emmagatzemen informació sobre les interaccions de les proteïnes.
- Bases de dades amb les publicacions d'estudis biomèdics, on es mostren i discuteixen els resultats d'una multitud de treballs experimentals.
- Bases de dades sobre vies de regulació de gens, és a dir, gens que sintetitzen proteïnes que activen a altres gens perquè passin a expressar-se.
- Bases de dades amb els gens classificats per la seva presumpta funció, la ubicació a la cèl·lula, etc.

El National Center for Biotechnology Information (NCBI) [12] forma part de la Biblioteca Nacional de Medicina dels Estats Units, una branca dels Instituts Nacionals de Salut. Es troba localitzat a Bethesda, Maryland i va ser fundat el 4 de novembre de 1988 amb la missió de ser una important font d'informació de biologia molecular. El NCBI conté les bases de dades anteriorment esmentades. Emmagatzema i constantment actualitza la informació referent a seqüències genòmiques en GenBank, un índex d'articles científics referents en els camps de la biomedicina, biotecnologia, bioquímica, genètica i genòmica a PubMed, una recopilació de malalties humanes i els gens implicats a l'OMIM, una classificació dels gens per funció i localització d'aquesta funció dins de la cèl·lula a GO, a més d'altres dades biotecnològiques de rellevància en diferents bases de dades.

Totes les bases de dades del NCBI [12] es troben disponibles online de manera gratuïta. Poden ser consultades via web utilitzant el buscador Entrez [12,14] via web i per aplicacions de tercers mitjançant les eines Eutils [15,16].

L'aplicació web del IBB [10] per trobar relacions gèniques a partir de bases de dades biomèdiques remotes [12-16] consta de les següents parts:

- L'applet java és l'encarregat de mostrar la xarxa de gens amb els gens de la microarray de l'usuari. Aquest applet permet seleccionar gens de la microarray i definir el tipus de relacions gèniques que es volen cercar.
- Aquest applet invoca un procediment que realitza el creuament de les bases de dades biomèdi-

ques i que mostra les relacions gèniques trobades per els paràmetres de la consulta.

Les dades que fa servir l'aplicatiu web que realitza el creuament de dades són obtingudes de tres fonts diferents:

- Bases de dades locals on es guarden les relacions gèniques entre gens de les bases de dades biomèdiques remotes. Aquestes bases de dades locals s'actualitzen periòdicament amb nova informació de les bases de dades remotes.
- Informació sobre la microarray que l'usuari està analitzant. Bàsicament el nombre dels gens.
- Consultes a les bases de dades remotes, principalment per obtenir informació sobre les relacions gèniques trobades a les bases de dades locals.

Les consultes a les bases de dades remotes del NCBI [12] es fan mitjançant les eines Eutils [15,16]. Utilitzant aquestes eines Eutils per consultar la base de dades PubMed obtenim tots els identificadors d'articles per un tòpic d'entrada. Això permet estudiar les relacions gèniques vinculades a una patologia o una funció cel·lular a l'aplicatiu actual.

Per desgràcia, l'aplicació web per cercar relacions gèniques amb gens de la microarray a partir de bases de dades biomèdiques no permet realitzar una cerca de relacions únicament a partir d'un tòpic d'entrada, sense entrar cap gen a relacionar. Aquesta funcionalitat podria ser molt útil per l'investigador o la investigadora, ja que si per exemple introdueix el nom d'una patologia podria conèixer els gens de la microarray relacionats amb aquesta patologia. Per exemple introduint *lipodistrofia*, que és l'efecte d'un fàrmac, podria conèixer tots els gens de la microarray que hi estan relacionats. No obstant, a l'hora de plantejar-se oferir aquesta nova funcionalitat, es presenten una sèrie de limitacions per part de l'aplicatiu.

Les consultes fent servir eines Eutils [15,16], en general ens retornen un gran nombre de resultats. Per exemple, si cerquem a la base de dades PubMed del NCBI el tòpic: *càncer*, la consulta ens retorna 3066855 resultats. L'aplicació actualment no està preparada per tractar aquesta gran quantitat de dades.

L'aplicatiu cerca relacions entre gens de la microarray amb altres gens, siguin o no a la microarray d'estudi. No obstant, en molts casos l'investigador o la investigadora voldria cercar relacions únicament entre gens de la microarray.

Lamentablement actualment l'aplicatiu web per cercar relacions gèniques només permet trobar relacions gèniques a partir dels gens seleccionats de la microarray de l'usuari. No deixa per exemple, introduir gens que no apareixen a la microarray per trobar quins gens de la microarray estarien relacionats amb aquests gens externs.

La introducció manual de gens resultaria una opció molt interessant per l'investigador o la investigadora.

Tenim un altre problema a l'aplicatiu web i és que no permet seleccionar un gran nombre de gens a la microarray de l'usuari. Aquest problema és degut a un problema de comunicació entre l'applet de Java que mostra la xarxa de gens i l'aplicatiu PHP que realitza el creuament de les bases de dades.

L'eina per cercar relacions gèniques actualment presenta també un altre problema i és que no permet a l'investigador o investigadora cercar relacions gèniques pels mateixos gens d'entrada però en diferents taxonomies. Això resultaria molt interessant, per exemple, per poder extrapolar resultats d'experimentació animal a humans.

Moltes relacions gèniques relacionen múltiples gens, per exemple un mateix article de la base de dades Pubmed pot parlar de múltiples gens, o bé una funció de GO. L'eina també pateix de carències a l'hora de permetre analitzar les diferents relacions entre gens i veure quins gens es veuen involucrats per les mateixes relacions. Aquesta funcionalitat resultaria molt útil per veure tots els gens que estarien involucrats en una mateixa relació o estudiar quins gens comparteixen múltiples relacions, és a dir, que mantenen les mateixes relacions gèniques amb els gens d'entrada. Com per exemple dos articles de la base de dades PubMed que relacionen els mateixos gens amb un gen d'entrada.

## 1.2 Objectius

L'objectiu principal que es planteja és fer més potents i eficients les consultes que realitzen els usuaris per trobar les relacions gèniques. Tot coneixent els problemes i carències que existeixen actualment en l'aplicatiu web per cercar relacions gèniques a partir de les bases de dades biomèdiques.

Per exposar els objectius del desenvolupament dividiré els objectius en subparts en funció de l'entorn funcional al qual afecten. Els objectius més rellevants del treball són aconseguir tractar la gran quantitat d'informació amb la que treballem i permetre cercar les relacions existents entre els gens de la microarray i un tòpic concret.

Objectius relacionats amb l'aplicatiu php que realitza l'encreuament entre les bases de dades biomèdiques locals i remotes:

1. Permetré la possibilitat de cercar gens relacionats amb un tòpic d'interès sense introduir gens d'entrada.
2. Aconseguiré treballar amb un gran volum de nombres de dades a les consultes a PubMed.
3. Oferiré la possibilitat de filtrar les relacions gèniques resultants mostrant només relacions gèniques entre gens de la microarray.

4. Oferiré la possibilitat d'estudiar les relacions gèniques dels gens d'entrada per a totes les taxonomies.
5. Permetré introduir manualment gens externs a la microarray per cercar les seves relacions gèniques amb els gens de la microarray. Aquests gens es podran introduir pel seu nom oficial o en qualsevol de les seves nomenclatures.
6. Tot l'esmentat s'ha de realitzar en un temps de resposta raonable donat que es tracta d'una aplicació on-line.

Objectius relacionats amb les operacions sobre els resultats mostrats (les relacions gèniques):

1. Habilitaré un sistema per seleccionar les relacions entre gens i mostrar tots els gens relacionats per una mateixa relació. Mostraré també els gens en comú entre diferents relacions gèniques (intersecció).
2. Milloraré les interfícies finals per fer-les més intuïtives i ràpides.

Objectius respecte a l'adaptació de l'applet que mostra la xarxa de gens de la microarray (PCOPGene):

1. Permetré el funcionament en l'aplicatiu PHP sense haver seleccionat cap gen de la xarxa de gens de la microarray.
2. Cercaré noves estratègies per passar un gran nombre de gens de la xarxa de gens de la microarray a l'aplicatiu PHP.
3. Reparació de bugs detectats al llarg de procés de desenvolupament.

## 2 Resultats i Discussió

A continuació exposaré com he assolit els diferents objectius i justificaré perquè les decisions preses són les millors per assolir-los.

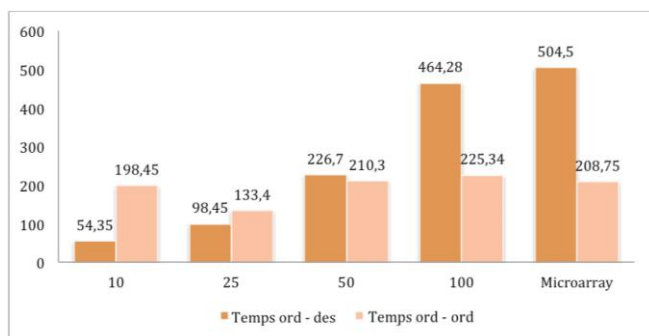
### **Aconseguir treballar amb un gran nombre d'identificadors d'articles de la base de dades PubMed per un tòpic d'interès.**

A partir de les bases de dades locals podem conèixer els articles vinculats als gens d'entrada, i a partir d'aquests articles, l'aplicació troba els gens relacionats amb els gens d'entrada per aparèixer a un mateix article. No obstant, quan volem només les relacions d'articles que parlen d'un tòpic concret, necessitem a més saber quins articles estan vinculats amb aquest tòpic. Com aquesta informació no és a la base de dades locals, cal consultar les bases de dades remotes per a que ens proporcionin tots els articles vinculats al tòpic que ha entrat l'usuari. Una vegada tenim la llista d'articles vinculats al tòpic d'entrada i la llista d'articles vinculats als gens d'entrada, hem de fer un *matching* per trobar els articles vinculats a l'hora als gens d'entrada i al tòpic d'entrada.

Per obtenir els identificadors d'articles de la base de dades remota, primer vam provar de convertir-los en un objecte XML i utilitzar les funcions natives de PHP per recorre'ls i fer el *matching*. Però això no és possible perquè els resultats obtinguts d'una consulta EUtils no poden ser tractats com un objecte XML. La segona alternativa passa per utilitzar les funcions predefinides de PHP de lectura de fitxers, la funció *fopen* i *fgets*.

Una vegada sabem com obtenir els articles vinculats al tòpic d'entrada s'ha de fer el *matching*. Tenint en compte que tenim els articles relacionats amb els gens d'entrada en una array, podríem utilitzar la funció *in\_array* de PHP per fer un *matching* on-line a mesura que llegim els articles del servidor remot. Aquesta solució, encara que ens evita emmagatzemar els identificadors d'articles obtinguts de les bases de dades remotes, és inviable si el nombre d'articles vinculats als gens d'entrada és massa gran, ja que estariem parlant d'un cost  $n*m$ .

La solució passa per fer un *matching* entre dos llistes ordenades. No obstant els identificadors obtinguts de la consulta a la base de dades biomèdica remota PubMed no es troben ordenats, i el PHP és incapaç d'emmagatzemar tanta informació i molt menys ordenar-la. En canvi, la llista d'identificadors d'articles vinculats als gens d'entrada si que pot ser ordenada amb la funció *sort* de PHP degut a que té un menor tamany. Per ordenar la llista d'articles de la base de dades remota descarreguem un fitxer amb els identificadors i l'ordenem mitjançant comandes Unix.



**Figura 1.** A la gràfica es pot veure el temps de diferencia d'utilitzar una llista d'articles ordenada i l'altre desordenada, o utilitzar dues llistes ordenades en el *matching* per cercar els articles vinculats als gens d'entrada que parlen d'un tòpic determinat. El temps (eix y) ve donat en segons. El label Microarray indica que s'entren tots els gens de la microarray de proves (un total de 1614). Utilitzar la llista d'articles vinculats als gens d'entrada ordenada però la llista d'articles que parlen d'un tòpic determinat desordenada, és més òptim quan el nombre de gens d'entrada és petit (10 i 25 a la figura). Si el nombre de gens d'entrada és gran ens convé ordenar també la llista d'articles que parlen d'un tòpic determinat.

D'aquesta manera s'habiliten dues opcions depenent del nombre d'articles vinculats als gens d'entrada: si aquest nombre és petit, ordenarem els articles vinculats als gens d'entrada i farem el *matching* on-line mentres llegim els articles del servidor remot; si el nombre d'articles vincu-

lats als gens d'entrada és massa gran, farem un *matching* amb les dues llistes ordenades. La figura 1, mostra el nombre d'articles vinculats als gens d'entrada on convé canviar d'estratègia. Els fitxers generats es creen amb la nomenclatura explicada a metodologia.

### Oferir la possibilitat d'estudiar les relacions gèniques dels gens d'entrada per a totes les taxonomies.

Quan l'usuari selecciona la opció d'estudiar les relacions dels gens d'entrada per a totes les taxonomies, els gens d'entrada passen a ser tots aquells que tenen el símbol del gen d'entrada en altres taxonomies diferents a la de la microarray analitzada. Disposem del vincle entre els símbols dels gens i les taxonomies a la base de dades local. D'aquesta forma, realitzo una consulta per obtenir un gen en totes les seves taxonomies al inici i després tracto el gen de cada taxonomia com si es tractés d'un gen diferent.

Per exemple el gen HSD3B1 és present en les següents espècies: *Danio rerio*, *Gallus gallus*, *Homo sapiens*, *Sus scrofa*, *Bos taurus*, *Mus musculus*, *Rattus norvegicus*, *Xenopus laevis*, *Macaca mulatta*, *Pan troglodytes*, *Canis lupus familiaris*, *Taeniopygia guttata*, *Oncorhynchus mykiss*, etc. Amb la nova funció implementada ens permetrà veure tots els gens relacionats de manera independent per cada espècie.

El punt més crític d'aquesta funcionalitat, ha sigut el tractament AND o d'intersecció entre els gens relacionats amb els gens d'entrada. Com que no té cap sentit buscar gens comuns entre diferents taxonomies, s'ha de fer una intersecció independent per a cada taxonomia.

### Oferir la possibilitat de filtrar les relacions gèniques resultants mostrant només relacions gèniques entre gens de la microarray.

Per mostrar relacions gèniques només entre gens de la microarray d'estudi, primer hem d'obtenir tots els gens de la microarray, i després fer un *matching* entre els gens relacionats amb els gens d'entrada i els gens de la microarray. Només els gens relacionats amb els gens d'entrada, que també siguin gens de la microarray, s'acabaran mostrant per pantalla. Degut a que es tracta de dues llistes d'identificadors no excessivament grans, el coll d'ampolla en aquest cas prové de l'obtenció dels gens de la microarray. Una possibilitat per obtenir els gens de la microarray és llegir la taula de la base de dades local on són tots els gens de totes les microarrays pujades al servidor(gene). Però també disposem d'aquesta informació a un fitxer (més detalls a la secció metodologia). Aquesta segona opció es la més adient doncs el fitxer conté només els gens de la microarray d'estudi. Una vegada tenim el símbol dels gens obtenim de la base de dades local el identificador oficial de cada gen i fem el *matching* amb els identificadors dels gens relacionats amb els gens d'entrada.

## Oferir la possibilitat de cercar gens vinculats a un tòpic d'interès per a tots els gens de la microarray.

Ara oferim a l'investigador o investigadora la possibilitat de conèixer els gens de la microarray d'estudi vinculats a un tòpic del seu interès. Per poder realitzar aquesta tasca, el programa obté els noms dels gens del fitxer utilitzat per només mostrar els gens de la microarray relacionats amb els gens d'entrada (secció anterior). A partir del símbol dels gens realitza una consulta a les taules de la base de dades locals (*gene\_info*) per obtenir l'identificador oficial de cada gen. Llavors tots aquests identificadors passen a ser els gens d'entrada i es realitza un tractament normal, excepte a l'hora de mostrar els resultats.

Per poder tractar amb un nombre tant gran de gens d'entrada i articles associats, una vegada tenim els gens i articles del *matching* per obtenir els articles relacionats amb els gens d'entrada i el tòpic d'entrada (primera subsecció de resultats i discussió) s'emmagatzemen en un diccionari PHP.

A aquest diccionari PHP la clau és l'identificador del gen d'entrada i els valors de cada clau són els identificadors d'articles PubMed relacionats amb el gen d'entrada. Aquest procés es realitza sempre que l'usuari vol realitzar una cerca de relacions gèniques en base a articles i per un tòpic d'entrada. D'aquesta forma obtenim tota la informació en una sola consulta a les bases de dades remotes del NCBI, i després podem tractar la informació obtinguda (després dels corresponents *matching* i post-tractament) de forma independent i amb un accés ràpid.

A l'hora de mostrar els resultats, ens interessa mostrar: 1) gens de la microarray que mantinguin una relació amb el gen d'entrada per mitjà d'un article vinculat al tòpic d'entrada; 2) la resta d'articles que relacionen el gen d'entrada i el tòpic d'interès, encara que no involucrin cap altre gen de la microarray. Aquestes dues llistes es mostren per a cada gen d'entrada (a no ser que estiguin buides). Tal com es pot veure a la figura 2.

BRMS1L	
Gene	PubMed
SMARCB1: SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1	ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome e
1 entry	1 Genes found

Extra PubMed	
Generation and Initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Identification of a novel BRMS1-homologue protein p40 as a component of the mSin3A/p33(ING1b)/HDAC1 hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes.	

**Figura 2.** S'ha realitzat una consulta per a articles vinculats al càncer sense gens d'entrada. A la imatge es mostra el resultat pel gen BRMS1L, d'entre els gens obtinguts. En la primera llista podem veure els gens que formen part de la microarray relacionats amb BRMS1L i l'article que els relaciona. En la segona llista mostra els articles relacionats amb BRMS1L, que parlen de càncer, però que no estan relacionats amb cap altre gen de la microarray.

Aquesta llista d'articles vinculats al gen i al tòpic d'entrada però a cap altre gen de la microarray s'obtenen quan es munta el diccionari PHP amb els articles relacionats amb cada gen d'entrada i el tòpic d'entrada

## Oferir la possibilitat d'introduir manualment gens externs a la microarray per cercar les seves relacions gèniques. Només introduint el seu nom oficial o en qualsevol de les seves nomenclatures.

La consulta manual per qualsevol nomenclatura d'un gen ha estat possible mitjançant una consulta a la base de dades local (*gene\_info*), on disposem d'un camp amb totes les possibles nomenclatures addicionals de cada gen. Primer es consulta si el text entrat manualment és un símbol de gen, si no es troba es consulta si és l'aliè d'un gen. Els gens obtinguts passen a ser gens d'entrada juntament amb els seleccionats a la xarxa de gens de la microarray. Tot el creuament de dades treballa amb els identificadors oficials i internacionals dels gens.

## Un cop mostrades les relacions gèniques trobades, mostrar els gens relacionats per una mateixa relació i els gens en comú entre diverses relacions.

Si ens fixem a la figura 3, sota el nom de cada gen relacionat se'ns mostra el nombre de relacions gèniques entre aquest gen relacionat i el gen d'entrada pels paràmetres de la consulta. Aquest número és un link que ens obre una nova finestra al navegador amb aquest llistat de relacions.

CASC3 TOPIC PUBMED = 'cancer' TAX: Homo sapiens CASC3	
Gene	PubMed
mag: mago nashi	An eIF4AIII-containi...
1 entry	7 Genes found
eIF4AIII: CG7483 gene product from transcript CG7483-RA	An eIF4AIII-containi...
1 entry	7 Genes found
osk: oskar	An eIF4AIII-containi...
1 entry	7 Genes found
btz: barentsz	An eIF4AIII-containi...
1 entry	7 Genes found
MAGOH: mago-nashi homolog, proliferation-associated (Drosophila)	An eIF4AIII-containi...
1 entry	7 Genes found
EIF4A3: eukaryotic translation initiation factor 4A3	An eIF4AIII-containi...
1 entry	7 Genes found
eIF-4a: Eukaryotic initiation factor 4a	An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay.
1 entry	7 Genes found

**Figura 3.** En la imatge podem veure alguns dels gens relacionats amb el gen CASC3 quan el tòpic d'entrada a la base de dades PubMed és càncer. Sota el nom del gen es mostren el nombre de relacions gèniques en que aquest gen relacionat participa. A la part dreta veiem els títols dels articles que relacionen els gens, i sota cada títol el nombre de gens relacionats per aquest article.

Seleccionant els checkbox com veiem a la figura 4, una funció AJAX captura les relacions seleccionades (en aquest cas articles) i llença una consulta a la base de da-

des local que ens retorna els gens relacionats per la relació seleccionada. En el cas de seleccionar-ne més de una es realitza una intersecció, obtenint els gens comuns entre tots els articles seleccionats.

CASC3  
TOPIC PUBMED = 'cancer'  
TAX: Homo sapiens  
CASC3

Gene	PubMed
mago: mago nashi	An eIF4AIII-containi... 7 Genes found
1 entry	
eIF4AIII: CG7483 gene product from transcript CG7483-RA	An eIF4AIII-containi... 7 Genes found
1 entry	
osk: oskar	An eIF4AIII-containi... 7 Genes found
1 entry	
btz: barentsz	An eIF4AIII-containi... 7 Genes found
1 entry	
MAGOH: mago-nashi homolog, proliferation-associated (Drosophila)	An eIF4AIII-containi... 7 Genes found
Genes result with de marked options	
MAGOH: mago-nashi homolog, proliferation-associated (Drosophila)	EIF4A3: eukaryotic translation initiation factor 4A3
eIF-4a: Eukaryotic initiation factor 4a	mago: mago nashi
eIF4AIII: CG7483 gene product from transcript CG7483-RA	osk: oskar
btz: barentsz	

**Figura 4.** S'ha seleccionat un article a la consulta de la figura 3. Això ha obert una finestra AJAX on apareixen tots els gens relacionats per aquest article amb els gens d'entrada. En cas de que es seleccionés més d'un article, el programa realitzaria una intersecció i només mostraria aquells que són comuns.

### Millora de les interfícies finals per fer-les més intuïtives i ràpides.

A la capçalera es mostren en tot moment els paràmetres seleccionats per l'usuari, així com els gens d'entrada. A les figures 5, 6 i 7 es mostren algunes de les optimitzacions que s'han realitzat.

TOPIC PUBMED = 'cancer'  
TAX: Homo sapiens  
ONLY RELATE GENES FROM EXPRESION MATRIX

**Figura 5.** Informació que li apareix a l'usuari abans de les llistes de gens relacionats. En aquest exemple s'ha realitzat una cerca sense gens on el títol de la base de dades biomèdica PubMed és càncer.

OR  
CASC3  
PERP  
TP53  
TOPIC PUBMED = 'cancer'  
TAX: Homo sapiens

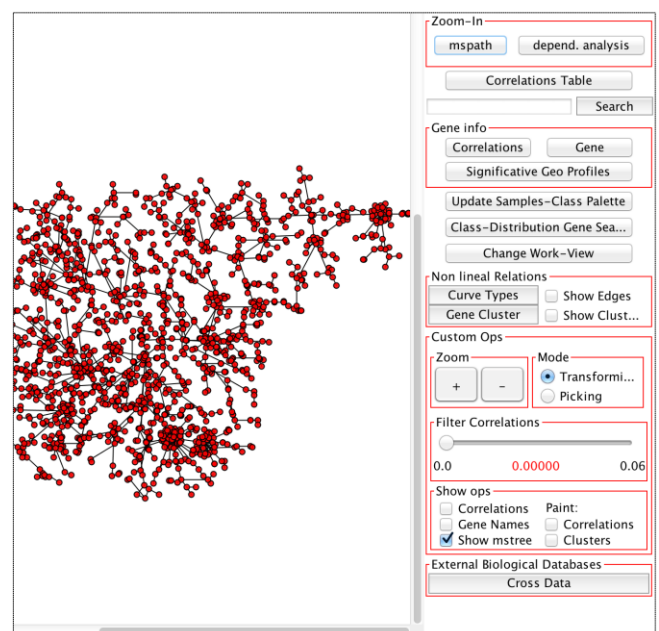
**Figura 6.** Informació que li apareix a l'usuari abans de les llistes de gens relacionats. En aquest exemple s'ha realitzat una cerca OR (unió de gens relacionats) on els gens d'entrada són CASC3, PERP, TP53.. El tòpic de la base de dades biomèdica remota PubMed és càncer. No s'ha activat la consulta per a totes les taxonomies per això s'indica que la taxonomia és la de la microarray.

PubMed
An eIF4AIII-containi... 7 Genes found
An eIF4AIII-containi... 7 Genes found
An eIF4AIII-containi... 7 Genes found
An eIF4AIII-containi... 7 Genes found
An eIF4AIII-containi... 7 Genes found
An eIF4AIII-containi... 7 Genes found
An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. 7 Genes found

**Figura 7.** Resultats d'una consulta per articles de la base de dades biomèdica PubMed. Al fer click sobre la capçalera PubMed a la part superior de la taula, el programa ordena els resultats pel títol de l'article. Per millorar la visibilitat s'ha decidit escorçar els títols d'articles que relacionen múltiples gens a 15 caràcters, per facilitar la visualització.

### Realitzar els processos en un temps de resposta raonable donat que es tracta d'una aplicació on-line

Un factor important en el nostre aplicatiu és el temps de resposta, donat que es una aplicació web en temps real i que s'està creuant tota la informació biomèdica coneguda. Els punts més crítics són els processos de *matching* entre identificadors.



**Figura 8.** Xarxa de gens que relaciona per expressió dels gens de la microarray analitzada. El botó *Cross Data* crida al PHP que provarà relacionar aquests gens entre ells i amb altres gens que no són a la microarray i "tòpics" d'interès utilitzant bases de dades biomèdiques internacionals.



Un altre punt crític que també complica el temps de resposta és que totes les operacions esmentades són combinables entre elles, així com amb la intersecció entre els gens relacionats a partir de les diferents bases de dades, i la intersecció entre els gens relacionats amb els diferents gens d'entrada (operació AND). D'aquests processos d'intersecció només s'obtenen els gens comuns a tots ells.

### Adaptacions al applet de Java que mostra la xarxa entre gens de la microarray.

En la figura 8 podem veure la xarxa de gens de la microarray de l'usuari. S'ha modificat l'Applet perquè es permeti realitzar el creuament sense haver fet click a cap gen. S'ha testejat tota la crida al PHP des de l'applet i s'han reparat els bugs existents.

**Figura 9.** Menú de selecció de paràmetres per cridar al PHP que fa el creuament amb les bases de dades biomèdiques. Source genes indica els gens de la microarray seleccionats a la xarxa de gens de la microarray. More Genes permet introduir noms de gens pel seu símbol o qualsevol dels seus noms coneguts. Crossing Parametres indica les bases de dades on es vol cercar gens relacionats amb els gens d'entrada. Si es selecciona Pubmed, Go Function, Go Process o Go Component, permet fer la cerca filtrada per un tòpic d'interès. Searching domains permet restringir el domini dels gens relacionats. Aquest poden ser només gens de la microarray o bé qualsevol gen existent. En la figura es mostra que no s'introdueix cap gen d'entrada, per tant, es farà una consulta de gens relacionats amb un tòpic, no de gens relacionats amb un altre gen.

### Cercar noves estratègies per passar un gran nombre de gens de la xarxa de gens de la microarray a l'aplicatiu PHP.

Si el nombre de gens a passar a l'aplicatiu PHP és prou gran, la URL del PHP que passa tots els paràmetres mitjançant el mètode GET de PHP es desborda.

La primera opció seria passar tots els paràmetres mitjançant un POST, però encara podríem tenir problemes de tamany.

La següent opció i finalment triada és crear un fitxer amb els gens seleccionats, si el nombre de gens d'entrada és

massa gran. No obstant l'applet no pot crear fitxers doncs per això cal que aquest estigui firmat.

La alternativa ha sigut passar els gens en parts a un PHP auxiliar que crea un fitxer que després el PHP de creuament de dades recupera i el tracta com el fitxer amb tots els gens de la microarray a la consulta sense gens d'entrada (secció anterior). Més detall a l'apartat de metodologia.

## 3 Metodologia

### 3.1 Directoris i Fitxers

Els fitxers amb els quals treballa l'aplicació per cercar les relacions gèniques es troben situats tant al servidor de desenvolupament com al servidor de producció del IBB, bàsicament en dos directoris.

#### Fitxers PHP

El fitxer PHP encarregat de rebre els gens, realitzar el creuament i mostrar els resultats es troba en el següent path:

`/var/www/html/applic/gexp/microarray/cruces/mostrar.php`

També hi ha dos fitxers PHP auxiliars a l'encarregat de rebre els gens, realitzar-ne el creuament i mostrar el resultat, que serveixen per ampliar la informació resultant:

`/var/www/html/applic/gexp/microarray/cruces/consulta.php`  
`/var/www/html/applic/gexp/microarray/cruces/allinfo.php`

#### Fitxers de l'Applet Java

L'applet Java es troba en la següent ruta:

`/var/www/html/applic/gexp/microarray/BioProject.jar`

Degut al gran tamany de les microarrays ens veiem obligats a particionar les xarxes de gens en diferents sectors, per això l'execució de l'applet és realitzada en una altra ruta en funció del sector que vulguem analitzar, per tant els fitxers que executarem tindran la següent nomenclatura:

`/var/www/html/applic/gexp/microarray/BioProjX.jar`  
 (On X és el número de la secció que volem analitzar)

#### Fitxers de dades

El fitxer que disposa de tots els gens de la microarray:

`/var/www/cgi-bin/pcop/microarray/mXX/XX.genes`  
 (On XX és el nombre de la microarray d'entrada).

#### Fitxers temporals

Tots els fitxers de dades temporals generats al llarg del procés són generats al directori:

`/var/www/html/applic/gexp/microarray/cruces`

Es generen un màxim de tres fitxers temporals durant l'execució.

Al inici del procés, l'applet Java es pot generar un fitxer, que bàsicament conté els noms dels gens d'entrada. El seu títol té el següent format:

*HH-mm-ss.txt*

On *HH* és la hora, *mm* els minuts i *ss* els segons.

Finalment en funció del nombre gens d'entrada el PHP *mostrar.php* pot generar dos fitxers que contindran identificadors d'articles de la base de dades biomèdica PubMed. El títol d'aquests fitxers tenen el següent títol:

*topic\_YYYY-MM-DD\_HH:mm:ss.txt*

*order\_YYYY-MM-DD\_HH:mm:ss.txt*

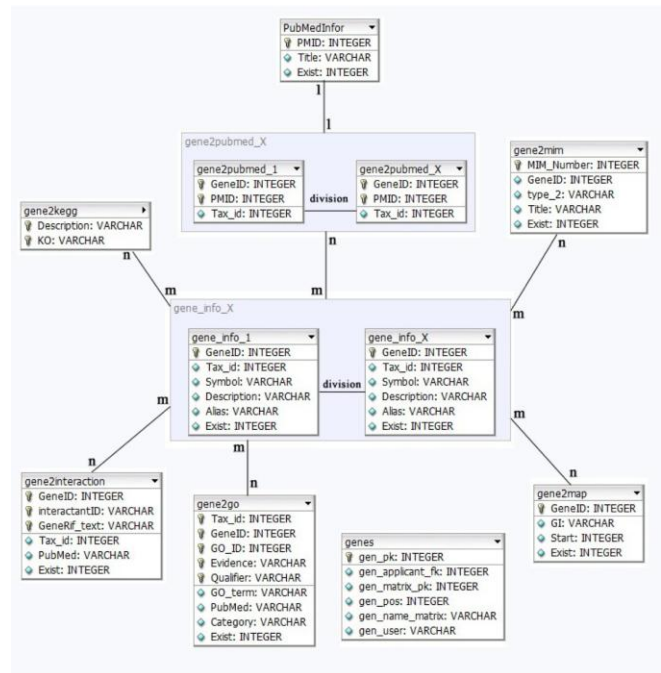
On *YYYY* és l'any actual, *MM* mes actual, *DD* dia actual, *HH* hora actual, *mm* minut actual i *ss* segons al instant de fer la petició.

### 3.2 Bases de dades

Les taules principals serveixen per obtenir informació dels gens, com per exemple, títol, descripció, taxonomia... etc i també per establir relacions entre els gens i en funció de les diferents bases de dades biomèdiques remotes.

#### Taules principals

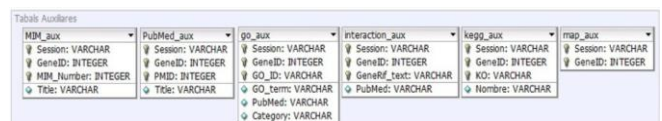
- *gene\_info\_X*: Taules que contenen tota la informació referent al gen. Id, taxonomia, símbol, descripció, alies (altres nomenclatures), etc..
- *gene2pubmed\_X*: Taules que relacionen els identificadors dels gens amb els identificadors dels articles de la base de dades PubMed del NCBI.
- *gene2mim*: Taula que relaciona els identificadors dels gens amb la base de dades MIM.
- *gene2map*: Taula que relaciona els identificadors dels gens amb la base de dades Map.
- *gene2go*: Taula que relaciona els identificadors dels gens amb la base de dades GO.
- *gene2interaction*: Taula que relaciona els identificadors dels gens amb la base de dades Interaction.
- *gene2kegg*: Taula que relaciona els identificadors dels gens amb la base de dades Kegg.
- *genes*: Taula que conté els noms actualitzats dels gens de cada microarray existent.
- *PubMedInfo*: Taula que conté els títols dels articles del base de dades PubMed del NCBI.



**Figura 10:** Estructura de la base de dades de l'aplicació web per cercar relacions gèniques. Creada sota un motor gestor de bases de dades *MYSQL*. En aquestes bases de dades és on es guarden les relacions gèniques entre gens de les bases de dades biomèdiques remotes. Aquestes bases de dades locals s'actualitzen periòdicament amb nova informació de les bases de dades remotes. També contenen informació sobre la microarray que l'usuari està analitzant. Bàsicament el nombre dels gens, en concret a les taules *gene\_info*.

#### Taules Auxiliars

Durant el procés la informació de creuament es va guardant en taules auxiliars que ens serveixen al final del procés per mostrar a l'usuari els resultats finals.



**Figura 11:** Estructura de les taules auxiliars de la base de dades de l'aplicació web per cercar relacions gèniques. Creada sota un motor de bases de dades *MYSQL*. Aquestes bases de dades s'utilitzen per anar emmagatzemant durant el procés d'obtenció de resultats, en aquest procés es van guardant tots els gens relacionats que es van trobant. Aquesta informació emmagatzemada és utilitzada al final del procés per mostrar-la a l'usuari per pantalla.

Com podem veure els noms d'aquestes taules auxiliars són anàlogues a les taules de la base de dades que contenen informació obtinguda de les bases de dades remotes. Disposem d'una taula auxiliar per cada base de dades remota.



### 3.3 Aplicació PHP

#### *mostrar.php*

Per cridar a aquest PHP que realitza el creuament de les dades, l'applet de java on es mostra la xarxa de gens de la microarray construeix una URL com aquesta: [http://revolutionresearch.uab.es/aplic/gexp/microarray/cruces/mostrar.php?opc=2&time=18-14-23&tax=9606&gene=ETS2&dom=1&microarray=17&umbral=0.1&topicfunc=&topiccomp=&topicproc=&use\\_tax=0&topicpub=&use\\_homo=0&logic=0&interactionOP=0&keggOP=0&pubmedOP=0&mimOP=0&mapOP=0&goFunctionOP=0&goComponentOP=0&goProcessOP=0&alias=Name1;Name2;...&nomFile=CDGS\\_20201517564609&](http://revolutionresearch.uab.es/aplic/gexp/microarray/cruces/mostrar.php?opc=2&time=18-14-23&tax=9606&gene=ETS2&dom=1&microarray=17&umbral=0.1&topicfunc=&topiccomp=&topicproc=&use_tax=0&topicpub=&use_homo=0&logic=0&interactionOP=0&keggOP=0&pubmedOP=0&mimOP=0&mapOP=0&goFunctionOP=0&goComponentOP=0&goProcessOP=0&alias=Name1;Name2;...&nomFile=CDGS_20201517564609&)

Paràmetres de la URL	
Opcions	Descripció
Tax	Indica la taxonomia del gen. Determina, junt al símbol, quin gen en concret està cercant l'usuari.
Gene	Símbol del gen. S'utilitza per determinar quin Gen vol tractar l'usuari junt amb la tax.
Dom	Identifica el domini dels gens que ha seleccionat l'usuari per restringir la cerca. Pot ser per el gens de la microarray o per tots els gens.
Microarray	Identificador de la microarray dels gens d'entrada. S'utilitza per poder identificar els gens relacionats que apareixen a la xarxa de gens.
Umbral	Defineix l'umbral de cerca en el cas d'homolegenes
topicfunc, topiccomp, topicproc	Paraules clau que es poden utilitzar per restringir la cerca per les bases de dades GO.
use_tax	Determina si l'usuari vol utilitzar la taxonomia de la microarray per la cerca o vol resultats per totes.
topicpub	Permet a l'usuari introduir les paraules clau per la cerca d'articles a Pubmed.
usehomo	Indica si utilitzem homogene per la cerca de gens.
logic	Indica quin tipus de creuament vol fer l'usuari pels resultats de cada gen d'entrada, AND o OR.
interactionOP, keggOP, pubmedOP, mimOP, mapOP, goFunctionOP, goComponentOP, goProcessOP	Indica la base de dades on l'usuari vol cercar gens relacionats amb els gens d'entrada
alias	Camp a través del qual enviem al programa PHP els gens extres que vol entrar l'usuari.
nomFile	Nom del fitxer on escriurem des de l'aplicatiu PHP els gens relacionats perquè l'applet Java els marqui a la xarxa de gens.

Taula 1. Opcions de la URL que crida al PHP principal.

El fitxer *mostrar.php* rep tots els paràmetres d'entrada que ha seleccionat l'usuari, els avalua, realitza els creuaments i mostra per pantalla els resultats a l'usuari.

En el cas de que es superin els 50 gens d'entrada i es tracti d'una consulta a la base de dades PubMed amb un tòpic d'entrada el procés crea un fitxer al mateix directori amb identificadors de PubMed amb la següent nomenclatura: *topic\_YYYY-MM-DD\_HH:mm:ss.txt*

On YYYY és l'any actual, MM mes actual, DD dia actual, HH hora actual, mm minut actual i ss segons al instant de fer la petició.

A continuació aquest PHP crea un altre fitxer idèntic però amb els identificadors dels articles de la base de dades PubMed ordenats fent servir la mateixa nomenclatura amb el següent format:

*order\_YYYY-MM-DD\_HH:mm:ss.txt*

Aquest PHP pot rebre un fitxer d'entrada en el cas en que existeixin un alt nombre de gens d'entrada degut a les limitacions del pas de noms de gens per paràmetre. Aquest fitxer per exemple, pot contenir tots els gens de la microarray seleccionada. Aquest fitxer de nom *HH-mm-ss.txt* (HH=hora, mm=minuts i ss= segons), en el cas de que es generi únicament conté els gens d'entrada que venen de l'applet Java.

El fitxer *mostrar.php* al llarg de la seva execució realitzarà consultes a totes les taules principals de la base de dades i també inserirà o actualitzarà les taules auxiliars.

#### *allinfo.php*

El fitxer *allinfo.php* rep l'identificador del gen a ser tractat i les dades de la sessió en que s'ha realitzat l'execució.

Tenint aquestes dades obté de les taules auxiliars tota la informació relacionada amb el gen d'entrada en aquella consulta a l'aplicatiu PHP. Aquesta informació són totes les relacions gèniques obtingudes d'aquell gen, per exemple, tots els articles de PubMed obtinguts al llarg del procés a partir de gen d'entrada.

#### *consulta.php*

El fitxer *consulta.php* és invocat des d'una funció Ajax del fitxer PHP *mostrar.php*. Aquest rep els identificadors de les entrades obtingudes per un determinat gen, per exemple els identificadors dels articles de la base de dades remotes de PubMed. També rep quina és la base de dades remota d'on s'ha obtingut aquesta relació gènica i les dades de la sessió en que l'usuari ha començat el procés de cercar de relacions gèniques. La funció Ajax del fitxer PHP *mostrar.php* permet seleccionar diferents relacions gèniques trobades i passar aquesta informació al fitxer *consulta.php*

Aquesta funció Ajax i amb aquesta informació que rep el fitxer *consulta.php*, mostra a l'usuari els gens relacionats mitjançant la relació gènica seleccionada i en cas de que es seleccionin més d'una relació gènica només mostra els

gens relacionats comuns. Això vol dir que només es mostren aquells gens que estiguin relacionats en totes i cadascuna de les relacions gèniques seleccionades.

Aquest PHP únicament realitza consultes a la taules auxiliars.

### 3.4 Applet Java PCOPGene

L'arxiu *BioProject.jar* és un arxiu que s'executa i que dins seu conté totes les classes i els *.java* necessaris pel funcionament de l'aplicatiu.

Hi ha noves funcionalitats en la classe principal *BioProjectPanel.java* degut als requeriments del desenvolupament. Un d'aquests canvis ha estat permetre l'accés al procés de creuament sense haver seleccionat cap gen de la microarray a la xarxa de gens.

D'altra banda, s'ha realitzat un canvi més crític que ha estat la creació d'un fitxer que conté noms de gens en el cas de tenir un alt nombre de gens a l'entrada. Aquest fitxer es crea al mateix directori on es troba el PHP i la seva nomenclatura és la següent:

*HH-mm-ss.txt*

On *HH* és la hora actual, *mm* el minut actual i *ss* els segons al instant de la petició.

### Llibreries JUNG

Les JUNG [17] són un framework de Java de codi obert que serveixen per el modelatge i la visualització. Aquestes llibreries porten una sèrie d'algoritmes de disseny incorporats que ens permeten mostrar les relacions gèniques entre gens de la microarray mitjançant un graf que en el seu conjunt construeix la xarxa de gens de la microarray.

## 4 Conclusions

Tal i com es detalla a continuació s'ha aconseguit assolir tots els objectius proposats al inici del treball:

Objectius relacionats amb l'aplicatiu php que realitza l'encreuament entre les bases de dades biomèdiques locals i remotes:

1. Es va permetre a l'aplicatiu la possibilitat de cercar gens relacionats amb un tòpic d'interès sense introduir gens d'entrada.
2. Es va trobar una solució pel tractament d'informació quan treballem amb un gran nombre d'identificadors.
3. S'ha aconseguit desenvolupar un procés de filtratge dels gens relacionats finals ràpid i eficaç, per només mostrar les relacions entre gens de la microarray de l'usuari.
4. Ara l'usuari a partir d'un gen d'entrada pot obtenir tots els resultats per totes les seves taxonomies.
5. A partir d'ara, l'aplicatiu permet a l'usuari introduir manualment gens externs a la microarray per cercar les seves relacions gèniques. Aquests

gens es poden introduir pel seu nom oficial o en qualsevol de les seves nomenclatures.

6. S'ha aconseguit un temps de resposta raonable, tant en el procés d'obtenció d'informació, el procés de creuament, com a la impressió del resultat final.

Objectius relacionats amb les operacions sobre els resultats mostrats (les relacions gèniques):

1. Es va habilitar un sistema per seleccionar les relacions entre gens i mostrar tots els gens relacionats per una mateixa relació. Mostra també els gens en comú entre diferents relacions gèniques (intersecció). El resultat es mostren una nova finestra carregada per Ajax cosa que ens permet el refresc instantani d'informació cada vegada que volem estudiar unes noves relacions gèniques amb els gens o tòpics d'entrada.
2. S'han controlat les peticions EUtils que realitza el programa contra el servidor NCBI per evitar interferències o la no recepció de tots els resultats.
3. S'han fet les adaptacions necessàries per tenir una interfície gràfica útil, intuïtiva i agradable per l'usuari.

Objectius respecte a l'adaptació de l'applet que mostra la xarxa de gens de la microarray (PCOPGene):

1. L'aplicatiu per cercar les relacions gèniques ara permet cercar les relacions sense haver seleccionat cap gen de la xarxa de gens de la microarray.
2. S'ha trobat una estratègia per passar un gran nombre de gens de la xarxa de gens de la microarray a l'aplicatiu PHP.
3. S'han detectat els bugs al applet i s'han solucionat.

El problema més complicat de solucionar ha estat que amb un volum d'informació tan gran les funcions PHP no estan preparades. Això ha portat a un treball d'investigació amb diferents estratègies, que a priori potser no semblaven les més atractives però que han acabat essent els més resolutives.

Al llarg del desenvolupament han anat sorgint molts problemes inesperats ja que és un procés on hi ha la intervenció d'un gran nombre de paràmetres, bases de dades, estructures de les dades, moltes fixades per entitats externes al IBB i que no es poden tocar. Tot i això, anar trobant la solució a tots aquests handicaps ha sigut un factor motivacional que ha assolit el seu pic al final de projecte.

## Bibliografia

- [1] Huerta M, Cedano J, Querol E. (2008): Analysis of nonlinear relations between expression profiles by the principal curves of oriented-points approach. J Bioinform Comput Biol.
- [2] Cedano J, Huerta M, Querol E. (2008): NCR-PCOPGene: An Exploratory Tool for Analysis of Sample-Classes Effect on Gene-Expression Relationships Advances in Bioinformatics, vol. 2008
- [3] Huerta M, Cedano J, Peña D, Rodriguez A, Querol E. (2009): PCOPGene-Net: holistic characterisation of cellular states from microarray data base on continuous and non-continuous analysis of gene-expression relationships. BMC Bioinformatics 2009 May 9;10:138.
- [4] Huerta M, Munyi M, Expósito D, Querol E, Cedano J.(2014) MGDB: crossing the marker genes of a user microarray with a database of public-microarrays marker genes. Bioinformatics. 2014 Jun 15;30(12):1780-1. doi: 10.1093/bioinformatics/btu109
- [5] Huerta M, Fernández-Márquez J, Cabello JL, Medrano A, Querol E, Cedano J.(2014) Analysis of gene expression for studying tumor progression: the case of glucocorticoid administration. Gene. 2014 Oct 1; 549(1):33-40. doi: 10.1016/j.gene.2014.07.027
- [6] Huerta M, Casanova O, Barchino R, Flores J, Querol E, Cedano J.(2014) Studying the complex expression dependences between sets of coexpressed genes. Biomed Res Int. 2014;2014:940821. doi: 10.1155/2014/940821.
- [7] Delicado, P.(2001): Another look at principal curves and surfaces. Journal of Multivariate Analysis, 77, 84-116 .
- [8] Delicado, P. and Huerta, M. (2003): 'Principal Curves of Oriented Points: Theoretical and computational improvements'. Computational Statistics 18, 293-315.
- [9] Cedano J, Huerta M, Estrada I, Ballosera F, Conchillo O, Delicado P, Querol E. (2007): A web server for automatic analysis and extraction of relevant biological knowledge. Comput Biol Med. 37:1672-1675.
- [10] Institut de Biotecnologia i de Biomedicina (IBB) de la Universitat Autònoma de Barcelona. <http://ibb.uab.es/ibb/>
- [11] WebServer per l'anàlisi online de MicroArrays del IBB <http://revolutionresearch.uab.es/>
- [12] Web oficial del National Center for Biotechnology Information (NCBI) <http://www.ncbi.nlm.nih.gov/>
- [13] KEGG (Kyoto Encyclopedia of Genes and Genomes) <http://www.genome.jp/>
- [14] Cercador web mitjançant el qual accedim a les bases de dades del NCBI. <http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>
- [15] Eines WebService gratuïtes del NCBI <http://www.ncbi.nlm.nih.gov/books/NBK25500/>
- [16] National Center for Biotechnology Information (US); 2010. Entrez Programming Utilities Help Bethesda (MD)
- [17] JUNG Java Universal Network/Graph Framework <http://jung.sourceforge.net/>